Prepared for the
Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting.
San Diego, CA, July 30 - August 4, 2000

METHODOLOGY FOR EVALUATING A SIMULATOR'S CAPABILITY TO SUPPORT ASSESSMENT OF MARINER PROFICIENCY

Mireille Raby
University of Iowa
Iowa City, Iowa

Alice M. Forsythe, Marvin C. McCallum
Battelle Seattle Research Center
Seattle, Washington

Myriam W. Smith
U.S. Coast Guard Research & Development Center
Groton, Connecticut

The goal of this project was to develop a structured process for evaluating the capability of simulators to support mariner performance assessment. In this approach, simulator evaluation objectives and criteria are based on performance-based mariner assessment requirements. The approach produces a simulator evaluation protocol that addresses four areas of simulator capability: (1) providing flexible exercise programming to the assessor; (2) replicating the characteristics of actual equipment; (3) simulating the operational conditions required to demonstrate the desired mariner performance; and (4) supporting thorough debriefing of the assessment candidate. A case study showed that our evaluation process was feasible. The process proved to be both sensitive and valuable in differentiating the capabilities of two simulators used to support mariner assessment objectives. This approach can be generalized to the evaluation of simulators of equipment used in other industries.

INTRODUCTION

Simulator evaluation is a complex issue for which no well-established procedure exists. Such evaluation often consists of examining a simulator's capability to replicate the functionality and/or dynamics of the actual equipment (e.g., its ability to replicate deceleration forces while braking in a vehicle). The limitation of this approach is that physical fidelity may not directly relate to functional or psychological fidelity. The utility of any evaluation depends on the purpose for which the simulator is to be used, and a simulator with high physical fidelity may not have the necessary capabilities and functionality to assess operator proficiency. This paper describes a process for evaluating the functional fidelity of simulators to support performance-based assessment (Raby, Forsythe, McCallum, & Smith, 1999).

The genesis of this project was the recent significant change in the United States Coast Guard (USCG) provisions for the training and qualifications of mariners. This change stemmed from the 1995 *Amendment to the Seafarer's Training, Certification and Watchkeeping (STCW) Code*, in which the International Maritime Organization (IMO) mandated that the assessment of mariner proficiency be done through practical demonstration to ensure competence for licensing (IMO, 1996). To be considered proficient, a mariner must now demonstrate the ability to perform shipboard operations safely and effectively either in a real environment (e.g., onboard a vessel) or in an operationally realistic setting.

Simulators provide a feasible alternative to demonstrating skills onboard an actual ship. A broad range of marine simulators is commercially available to support mariner assessment. Full-mission simulators commonly offer a highly realistic operational environment, consisting of a full-scale mock-up of the environment simulated combined with actual equipment. With the advances in personal computer (PC) processing capabilities, however, the distinction between PC-based desktop simulators and full-mission simulators is diminishing. Desktop simulators now simulate environments and equipment capabilities that are increasingly similar to those generated by more complex and expensive simulators. When these advances are considered in conjunction with the affordability of PC-based simulators, it is apparent that the capability of these desktop simulators must be seriously considered.

The primary objective of our work was to develop a systematic process for evaluating a simulator's capability to support specific mariner assessment procedures. This paper summarizes the step-by-step methodology we used to identify simulator requirements for performance-based assessment, to design a simulator evaluation protocol, and to test this methodology on two PC-based marine simulators.

METHODOLOGY

Figure 1 depicts the method used to assess the functional fidelity of the PC-based simulators. It is a five-step process. In the first step, performance-based mariner assessment requirements are defined. The requirements are based on an analysis of the mariner assessment objectives. Assessment objectives are identified through a review of mariner skill and knowledge requirements in specified operational areas. Basic mariner assessment requirements include assessment conditions, performance measures, and performance standards.

In the second step, simulator evaluation objectives (i.e., the specific items on which the simulator is evaluated), simulator evaluation conditions, and evaluation criteria are defined. All three are derived from the performance-based assessment requirements defined in the first step. The evaluation objectives are divided into four general categories of simulator capability: (1) providing flexible exercise programming to the assessor; (2) replicating the characteristics of actual equipment; (3) simulating the operational conditions required to demonstrate the desired operator performance; and (4) supporting thorough debriefing of the assessment candidate.

In the third step, the evaluation protocol is developed. The protocol consists of instructions and an evaluation form listing the evaluation conditions and criteria. The form is organized according to the simulator evaluation objectives and is divided into four categories: exercise programming, equipment set-up, simulation, and debriefing.

In the fourth step, the simulator evaluation is conducted. A separate evaluation should be conducted by at least two evaluators to ensure the results are reliable. In the fifth and final step, the findings of the evaluation are summarized and analyzed. It is noteworthy that this methodology can be applied to simulators of varying capabilities and cost, and it can be used to identify the most least costly simulator available to support the performance-based operator assessment requirements.

To provide a case study for the application of this methodology, we selected two PC-based simulators that reproduce the operation of an Automatic Radar Plotting

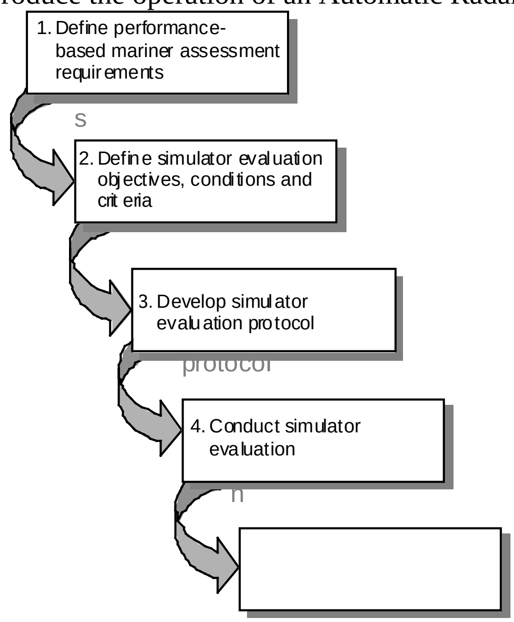

Figure 1. Proposed methodology for evaluating simulators used in performance-based assessments

Aid (ARPA), which is a piece of equipment used by members of a ship's bridge crew. As a basis for the ARPA simulator evaluation objectives, we analyzed the mariner performance assessment objectives and conditions defined for ARPA operation by McCallum, Forsythe, Smith, Nunnenkamp, & Sandberg (1999). We also investigated the characteristics of ARPA simulators required for valid assessment of mariner proficiency (Bole & Dineley, 1990; IMO, 1971, 1979, 1996). This analysis and research resulted in a list of 33 simulator evaluation objectives in four categories (exercise programming, equipment set-up, simulation, and debriefing). Each of the evaluation objectives had between two and 15 separate evaluation criteria associated with it (Raby, et al. 1999). The criteria represented specific requirements in control and display elements of the simulator.

We organized the evaluation objectives, conditions, and criteria into an evaluation protocol and then applied it to two commercially available desktop ARPA simulators in separate evaluations. We selected simulators with different processing characteristics and costs to ensure our evaluation protocol was sufficiently flexible for application to a range of simulators, and sufficiently sensitive to discriminate among them. The simulators are referred to as *X* and *Y*.

RESULTS

The simulators were rated on their ability to meet the criteria associated with each simulator evaluation objective. The rating process consisted of assigning a value of 1 (*yes*), 0.5 (*partial*), or 0 (*no*) to each evaluation criterion. Simulator scores for each objective were then calculated by summing the criteria ratings. The resulting scores identified the relative strengths and weaknesses of each simulator.

As an example, Table 1 compares the scores for simulators *X* and *Y* on each of the evaluation criteria corresponding to objective 2.1, *Selection of display presentation, orientation, and vector mode*. As the table indicates, simulator *X* fully met six criteria, partially met three criteria, and did not meet two criteria, resulting in a score of 7.5 for evaluation objective 2.1. In comparison, simulator *Y* fully met nine of the criteria and did not meet two of the criteria, resulting in a score of 9.0 for this evaluation objective.

Table 1. Summary of simulator capabilities for evaluation objective 2.1, *Selection of display presentation, orientation, and vector mode*.

Simulator Evaluation Criterion (C = Control, D = Display)		Evaluation Criterion Met: ***Simulator X***	Evaluation Criterion Met: ***Simulator Y***
2.1.C1	Ability to toggle between sea- and ground-stabilized modes	No	No
2.1.D1	Indication of display mode	No	No
2.1.C2	Ability to toggle between North-up, and either course-up or head-up azimuth stabilization	Partial	Yes
2.1.D2	Indication of display orientation mode	Yes	Yes
2.1.C3	Ability to toggle between relative and true motion	Yes	Yes
2.1.D3	Indication of display vector mode	Partial	Yes
2.1.C4	Ability to use ARPA on the following ranges: (a) 3 or 4 miles, and (b) 12 or 16 miles	Yes	Yes
2.1.C5	Fixed range rings available	Yes	Yes
2.1.D5.1	Indication of range scale in use	Yes	Yes
2.1.D5.2	Indication of distance between range rings	Partial	Yes
2.1.C6	Variable range marker available	Yes	Yes
Summary Score (Yes = 1, Partial = 0.5, No = 0)		7.5	9.0

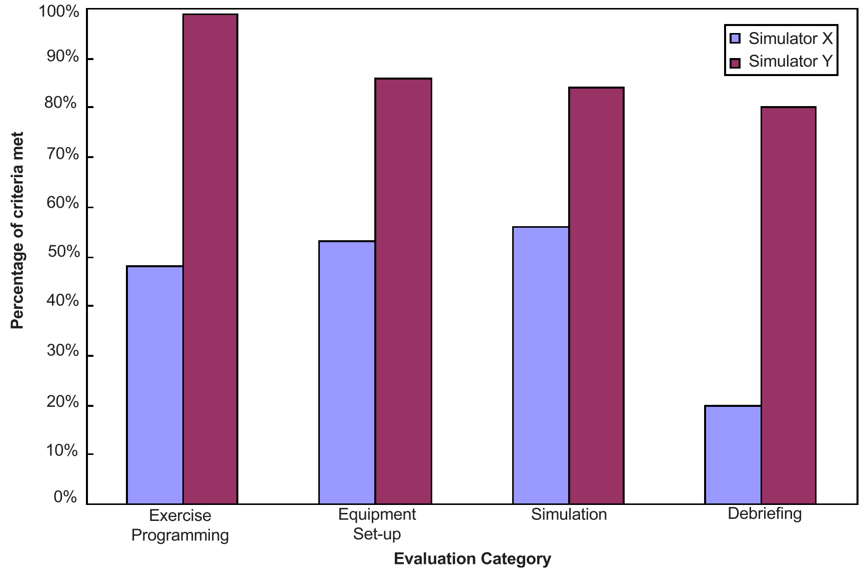

Figure 2. Percentage of criteria met by each ARPA simulator in four general simulator evaluation categories.

This scoring approach is quite simple. A more advanced approach would be to apply different weights to the individual simulator evaluation criteria prior to calculating summary scores. Higher weights would indicate those criteria that are considered relatively more important than other criteria. To obtain a weighted score, criterion scores would simply be multiplied by their respective weight. Valid criterion weights could be obtained through structured subject matter expert interviews.

To obtain a summary of the simulator's capabilities in supporting each of the four general simulator evaluation categories, scores for the individual evaluation objectives can be summed within these four evaluation categories. Figure 2 compares the two simulators evaluated in our case study with regards to the percentage of criteria met in the four general simulator evaluation categories. This figure reveals consistently high percentage scores for simulator *Y* (between 80 and 98 percent of the criteria met for each evaluation category), and more varied, but consistently lower percentage scores for simulator *X* (between 20 and 56 percent of the criteria were met). For example, with regards to *exercise programming*, simulator X could simulate landmasses and environmental conditions, but did not provide flexibility in specifying either the strength or weakness of the conditions. Comparatively, simulator Y had the ability to generate complex and varied exercise conditions and also to record the exercises for future use.

A simulator's capability to support operator assessment objectives is also a fundamental consideration that can be examined by compiling simulator evaluation objective scores corresponding to each of the operator assessment objectives. In our case study, simulator *Y* offered better overall support for mariner assessment, as it obtained consistently high percentage scores (between 83 percent and 100 percent) in supporting the six assessment objective categories specified in McCallum et al. (1999).

CONCLUSIONS

The project demonstrated the feasibility of the present approach to simulator evaluation. The approach is technically based on: (1) operator performance requirements; (2) assessment conditions required for demonstrating performance; and (3) operational requirements for the actual equipment used by the operator.

The ARPA simulator case study demonstrates the sensitivity and value of the method by identifying the differences between a more capable and higher-cost desktop simulator and a less capable, lower-cost one. The more capable, higher-cost simulator was able to support more of the assessment objectives, including those that required the candidate to demonstrate an understanding of the limitations of real equipment.

The present application of the simulator evaluation method was limited to PC-based ARPA simulators. However, the method is generic and has a broad range of potential applications, such as more complex ARPA simulators, other maritime simulator applications, or simulators designed for assessment of performance in other industries (e.g., flight simulators or driving simulators). This systematic approach to evaluating the functional fidelity of simulators represents an important tool for matching simulator capability to operator training and assessment requirements.

ACKNOWLEDGMENTS

The United States Coast Guard Research and Development Center funded this project under contract number DTCG39-94-D-E00777, delivery order number 96-F-E0014. The views expressed in this paper are those of the authors and are not official U.S. Coast Guard policy.

REFERENCES

Bole, A. G., & Dineley, W. O. (1990). *Radar and ARPA Manual.* Oxford, England: Butterworth-Heinemann.

International Maritime Organization. (1971). *Performance Standards for Navigational Radar Equipment* (Resolution No. A.222VII). London: Author.

International Maritime Organization. (1979). *Performance Standards for Automatic Radar Plotting Aids (ARPA)* (Resolution No. A.422XI). London: Author.

International Maritime Organization. (1996). *Seafarers' Training, Certification and Watchkeeping (STCW) Code.* London: Author.

McCallum, M. C., Forsythe, A. M., Smith, M. W., Nunnenkamp, J. M., & Sandberg, G. R., (1999). *Developing Performance-based Assessments of Mariner Proficiency* (Report No. RDC-303-99). Groton, CT: U.S. Coast Guard Research and Development Center.

Raby, M., Forsythe, A.M., McCallum, M.C., & Smith, M.W. (1999). *Evaluating Simulators for Assessment of Mariner Proficiency*. (Report RDC–150–99, in revision). Groton, CT: U.S. Coast Guard Research and Development Center.